ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Statistical modelling of uncertainty in satellite retrieval

Marko Laine

marko.laine@fmi.fi

Finnish Meteorological Institute

AURA Science meeting 2011-09-15

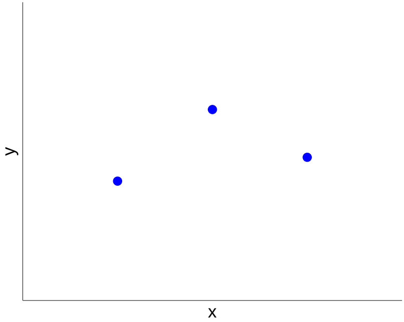
Statistical uncertainty

- Uncertainty in data
  - instrument noise
  - L1 data cleanup
  - L1 to L2 inversion
- Uncertainty in modelling
  - forward model formulation
  - *model selection*
  - representativeness
  - resolution
- Uncertainty in statistical inference
  - estimated model parameters
  - parameter uncertainty
  - model prediction
  - *interpolation*

Model selection

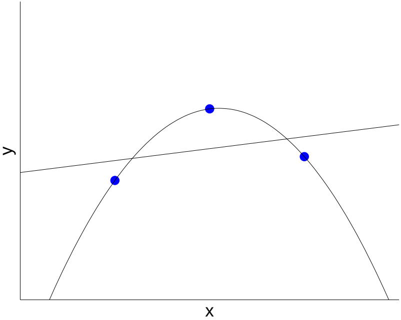
- Classical model parsimony problem
- AURA/OMI aerosol model selection for multi wavelength OMAERO retrieval algorithm.

Classical model parsimonity problem



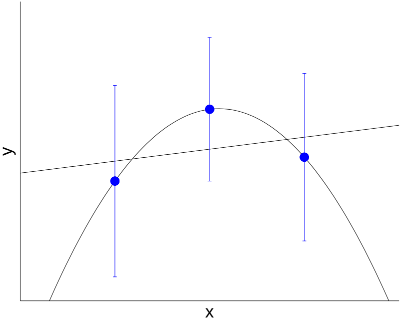
Classical model parsimonity problem

- Observations with two competing models
- Linear model with 2 parameters, quadratic with 3.



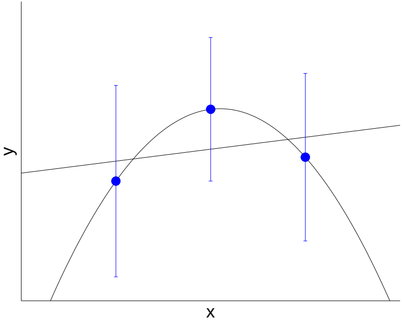
Classical model parsimony problem

- Observations with two competing models
- Linear model with 2 parameters, quadratic with 3.
- Only ofter having information on observation accuracy we can decide the goodness-of-fit.



Classical model parsimony problem

- Observations with two competing models
- Linear model with 2 parameters, quadratic with 3.
- Only ofter having information on observation accuracy we can decide the goodness-of-fit.
- How about bias?



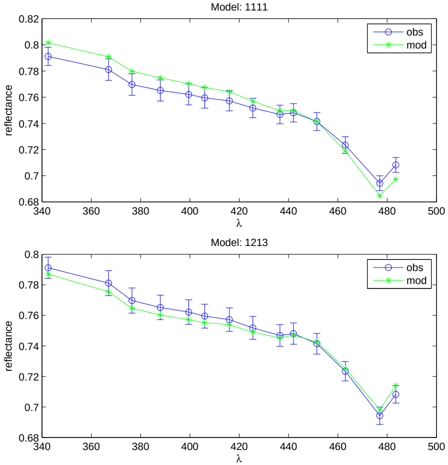
Bias and noise

In practice we need to take into account model discrepancy.

$$\text{obs} = \text{model} + \epsilon_{\text{noise}} + \epsilon_{\text{bias}}$$

Two aerosol models fitted to same AURA/OMI reflectance observations. Both fit the data within the limits of the accuracies of the observations given by error bars. Both also exhibit similar but opposite bias.

See Anu Määttä's poster.

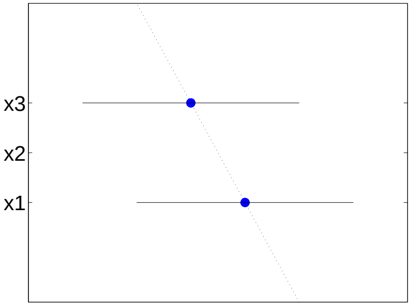


Beyond L2

- Profile interpolation
- Spatial interpolation
- Time series analysis
- Assimilation into atmospheric models

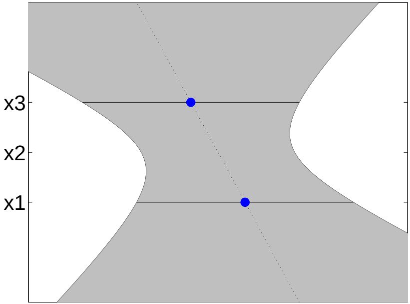
The problem of profile interpolation

- How to interpolate between two retrieved profile values?



The problem of profile interpolation

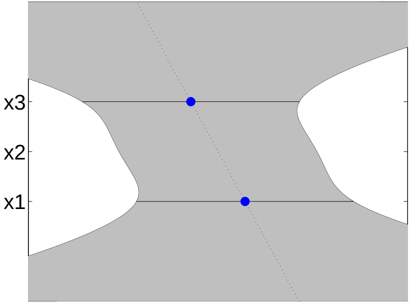
- How to interpolate between two retrieved profile values?
- Suppose the value in the middle is calculated as an average of the values above and below.



Setting $(x_1 + x_3)/2 - x_2 \sim N(0, s^2)$, with $s = 0$ means that we set – a priori – the second derivative of the profile to zero. This leads to standard error of mean as uncertainty for x_2.

The problem of profile interpolation

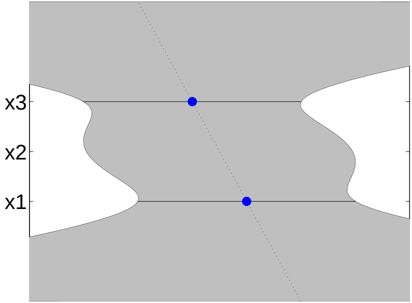
- How to interpolate between two retrieved profile values?
- Suppose the value in the middle is calculated as an average of the values above and below.



Setting $(x_1 + x_3)/2 - x_2 \sim N(0, s^2)$, with s equal to standard error of mean implies linear interpolation of uncertainty for x_2.

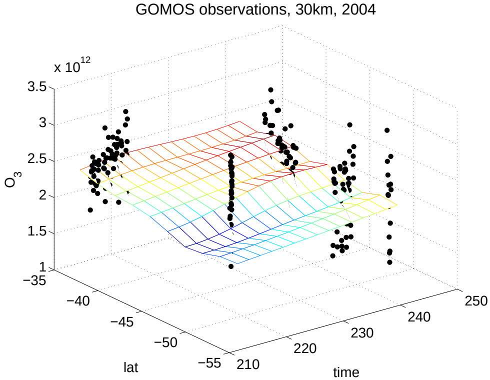
The problem of profile interpolation

- How to interpolate between two retrieved profile values?
- Suppose the value in the middle is calculated as an average of the values above and below.



Setting $(x_1 + x_3)/2 - x_2 \sim N(0, s^2)$, with s equal to mean of the individual uncertainties of x_1 and x_3 imply uncertainty as in linear regression when predicting a new observation.

Spatial interpolation for gridded data



- Same problem as in profile interpolation but in 3 dimensional LAT-LON-TIME grid.
- Griding non-uniform observations in space and time.
- A priori spatial variability (structure function) information needed.
- Can be implemented as hierarchical Gaussian model for predefined regular grid η.

Spatial interpolation using hierarchical Gaussian model

General model for spatial interpolation:

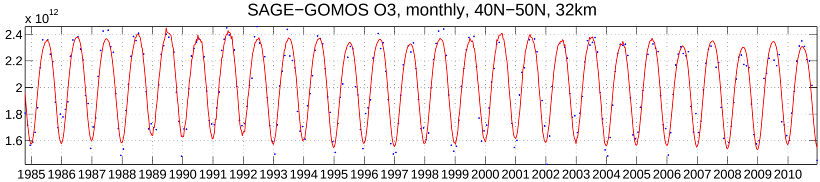
$$\begin{aligned} y|\eta &\sim N(H\eta, \Sigma_y), && \textit{observations} \\ \eta|\beta &\sim N(X\beta, \Sigma_\eta), && \textit{spatial random field} \\ \beta &\sim N(\beta_0, \Sigma_\beta) && \textit{hyper parameters} \end{aligned}$$

Notation: y observations, η spatial-temporal grid, H observation operator, X linear model for mean variability of O_3, β model parameters.

Determining realistic spatial correlation structure Σ_η is the most demanding task in the interpolation.

From gridded data to time series analysis

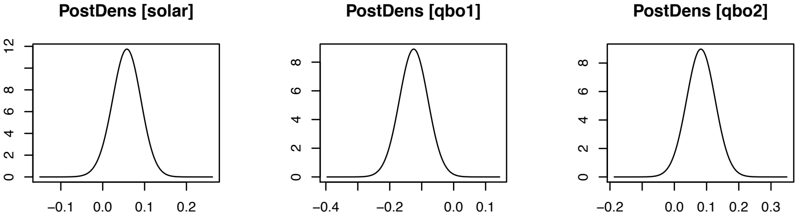
- Classical statistical methods usually not directly applicable to atmospheric time series.
- Preferable methods are those that directly apply to non-stationary series, using of proxies, and can handle observation uncertainty and missing observations, such as state space methods.
- Erkki Kyrölä's presentation.



Time series modelling

State space model for O_3 time series.

$$
\begin{aligned}
y_t &= x_t + \epsilon_t, && \textit{observations} \\
x_t &= \beta_1\, \mathrm{solar}(t) + \beta_2\, \mathrm{qbo1}(t) + \beta_3\, \mathrm{qbo2}(t) + \mu_t + s_t, && \textit{model} \\
\mu_t &= \mu_{t-1} + a_{t-1} + \epsilon_\mu, && \textit{local level} \\
a_{t-1} &= a_{t-2} + \epsilon_a, && \textit{local trend} \\
s_t &= -s_{t-1} - s_{t-2} - \ldots s_{t-11} + \epsilon_s, && \textit{seasonal effect}
\end{aligned}
$$



Conclusions and general thoughts

- Explicit and transparent a priori information means verifiable uncertainty statements.
- It allows comparison of a priori profiles to inverted profiles.
- Uncertainty estimates are always conditional to the assumptions about error statistics.
- This includes assumptions on the smoothness of underlying continuous fields.
- Model uncertainty analyses are usually heavily affected by the way model bias is handled.

Thank you

- H. Haario, M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen: MCMC methods for high dimensional inversion in remote sensing, *Journal of the Royal Statistical Society, Series B*, **66(3)**, pages 591–607, 2004. doi:10.1111/j.1467-9868.2004.02053.x
- M. Laine and J. Tamminen: Aerosol model selection and uncertainty modelling by adaptive MCMC technique, *Atmospheric Chemistry and Physics*, **8(24)**, pages 7697–7707, 2008. www.atmos-chem-phys.net/8/7697/2008/